Elements of secondary structure in a human epithelial mucin core peptide fragment

Saul J. B. TENDLER

Department of Pharmaceutical Sciences, University of Nottingham, Nottingham NG7 2RD, U.K.

The protein core of human epithelial mucin has previously been shown to consist of tandem repeats of a 20-amino-acid sequence that carries the epitopes for a number of tumour-marking monoclonal antibodies. High-field n.m.r. studies have now been undertaken on an 11-amino-acid fragment of this sequence dissolved in dimethyl sulphoxide. The studies reveal elements of secondary structure to be present: a type I β -turn has been identified from Asp² to Arg⁴ of this peptide, and this turn is extended by Pro⁵ being in the *trans* form. The observed turn region extends into the known epitopes for the antibodies C595 and NCRC-11 and may form the basis for how the antibodies recognize these peptides.

INTRODUCTION

Human epithelial mucins are complex glycoproteins of high M_r (400 000) that have been identified as the target antigens for many monoclonal antibodies produced against breast-carcinoma cells or milk-fat-globule membranes. They have a clinical relevance to breast cancer, since they are detectable in patients' serum at elevated levels where metastatic disease is present (Price, 1988). It has recently been established that the protein core of the epithelial mucin is composed of highly conserved tandem repeats of a 20-amino-acid sequence (peptide I in Table 1) (Gendler *et al.*, 1988). Price *et al.* (1990a) have identified many monoclonal antibodies that are able to recognize epitopes in synthetically prepared peptides bearing this sequence. A model for the secondary structure of the protein core has been developed on the basis of secondary-structure calculations (Price *et al.*, 1990b); this model is dominated by a hydrophilic-turn region extending from Pro¹ to approximately Thr¹⁰. This hydrophilic region has been shown by epitope-mapping experiments to contain the four-amino-acid (Arg-Pro-Ala-Pro) epitope for the antibody C595 and the three-amino-acid (Arg-Pro-Ala) epitope for NCRC-11 (Price *et al.*, 1990b).

High-field n.m.r. spectroscopy and, in particular, modern two-dimensional n.m.r. experiments, provide an elegant technique for the study of peptides and proteins in solution (Wuthrich, 1986). Although this technique has proved to be useful in the study of proteins, attempts to find structure in peptide fragments have not been extremely successful. This is probably due to many of the fragments under study having no regular folded conformation in solution on an n.m.r. timescale (Wright *et al.*, 1988). Antigenic

peptides, however, may exist in regular folded conformations which are selected to bind to the antibody and may therefore be studied by n.m.r. spectroscopy (Dyson *et al.*, 1986, 1988).

In order to elucidate the molecular-recognition events involved in the interactions between the mucin and specific monoclonal antibodies, it is first necessary to understand the solution structure of the glycoprotein by studying the antigenic peptide fragments. Preliminary n.m.r. temperature studies investigating the resonance position of the high-field methyl resonances of the full 20-amino-acid peptide I (Table 1) were not able to identify secondary structure in the peptide when it was dissolved in either water or trifluoroethanol (Price *et al.*, 1990b). The present study investigates the secondary structure of peptide II [p(1–11)] (Table 10), which has the sequence of the hydrophilic-turn region of the antigenic human epithelial mucin core peptide. The peptide dissolved in dimethyl sulphoxide has been studied by using a combination of one- and two-dimensional n.m.r. experiments.

EXPERIMENTAL

The peptide p(1–11) was prepared on an automated Milligen 9050 Pepsynthesiser and was purified by reverse-phase h.p.l.c. using a 0.1% trifluoroacetic acid/90% acetonitrile solvent gradient system. Samples for n.m.r. spectroscopy contained 3.96–4.67 mM-peptide and 0.7 mM-tetramethylsilane in [²H₆]dimethyl sulphoxide. ¹H n.m.r. spectra were recorded at 400 and 500 MHz on Bruker AM400 and AM500 n.m.r. spectrometers respectively, at temperatures from 293 to 313 K. Chemical-shift (δ) values (accurate to 0.01 p.p.m.) are reported relative to

Table 1. Sequences of peptides related to human epithelial mucin core protein

The one-letter notation for amino acids is used.

Peptide	Residue no. ...	Sequence				
		1	5	10	15	20
I		PDTRPAPGSTAPPA HGV TSA				
II [p(1–11)]		PDTRPAPGSTA				

Abbreviations used: COSY, correlation spectroscopy; HOHAHA, homonuclear-Hartmann-Hahn; NOESY, nuclear-Overhauser-enhancement spectroscopy; ROESY, rotating-frame Overhauser-enhancement spectroscopy.

Fig. 1. ^1H -n.m.r. spectrum of the peptide p(1-11)

The 400 MHz n.m.r. spectrum of the peptide dissolved in $[\text{D}_6]\text{dimethyl sulphoxide}$ was recorded at 313 K. The resonance at 2.5 p.p.m. corresponds to the residual protons in the solvent; the residual water peak was suppressed by pre-saturation.

Fig. 2. Amide-to-aliphatic region of the ^1H HOHAHA spectrum of the peptide p(1-11)

The 500 MHz ^1H HOHAHA spectrum of the peptide p(1-11) dissolved in $[\text{D}_6]\text{dimethyl sulphoxide}$ was recorded at 303 K. ω_1 and ω_2 are frequencies. The capital letters with the superscript numbers are amino acid residues given in the one-letter notation, the position in the sequence being given by the superscript number.

internal tetramethylsilane (δ 0.00 p.p.m.). One-dimensional spectra were collected in 16384 data points and routinely zero-filled to 32768 points to give a final resolution of 0.492 Hz/point. The two-dimensional correlation spectroscopy (COSY) experiments were recorded and analysed as described by Hammond *et al.* (1987). The homonuclear-Hartmann-Hahn-spectroscopy (HOHAHA) experiment (Braunschweiler & Ernst, 1983) was undertaken at 303 K with a spin locking time of 120 ms with 2048 data points recorded for each of the 300 increments. The rotating-frame Overhauser-enhancement-spectroscopy (ROESY) experiment (Bax & Davis, 1985) was performed at 303 K with a spin locking time of 200 ms, and 2048 data points were collected for each of the 300 increments. The spin locking time used in this experiment did not produce any problems associated with spin diffusion.

Fig. 3. ^1H ROESY spectrum of the peptide p(1-11)

^1H ROESY spectrum of the peptide p(1-11) dissolved in $[\text{D}_6]\text{dimethyl sulphoxide}$ was recorded at 500 MHz and at 303 K. The capital letters with superscript numerals, and ω_1 and ω_2 are defined in Fig. 2.

RESULTS AND DISCUSSION

The resonances of the peptide p(1-11) dissolved in dimethyl sulphoxide were initially assigned on the basis of chemical shift and connectivity in one-dimensional and two-dimensional COSY and HOHAHA experiments. The one-dimensional spectrum of the peptide, as shown in Fig. 1, demonstrates the general lack of degeneracy in this system. The amide-to-aliphatic region of the HOHAHA experiment giving the connections for the residues is presented in Fig. 2. The specific assignments of the residues were based on the use of the ROESY experiment (Fig. 3), as few through-space connections were observed in nuclear-Overhauser-enhancement (NOESY) experiments, owing to the tumbling rate of the peptide. The complete assignments for the peptide at 303 K are presented in Table 2. Pro¹ was assigned on the basis that it is the only proline residue present to have an amino proton. The two threonine residues were more difficult to assign specifically, especially as they have degenerate amide proton resonances at this temperature. Thr³ was, however, assigned on the basis of a through-space connection between its α -proton and the amide proton of Asp²; this assignment was confirmed by the presence of a through-space connection between the amide proton of Ser⁹ and the α -proton of Thr¹⁰. Pro⁵ was assigned on the basis of ROESY cross peaks between its two δ -proton resonances and the Arg⁴ amide proton resonance. A through-space connection between the amide proton of Gly⁸ and the α -proton of Pro⁷ made the assignment of the proline residues complete. All of the amino-acid-residue protons resonate at frequencies that are consistent with their type. It is, however, noteworthy that the amide proton of Asp² resonates at a particularly low-field position, which is suggestive of it being involved in hydrogen-bonding.

The temperature-dependence of the amide proton resonances of the peptide have been studied and are presented in Fig. 4. The low-field Asp² amide proton resonance has the lowest rate of movement observed in the system, with a temperature-dependence of 0.00287 p.p.m./K. The amide proton for Ser⁹ also appears to have a low rate of exchange. This indicates that these protons are shielded from the solvent, probably due to hydrogen-bonding in a β -type structure. The temperature-dependence for

Table 2. Chemical shifts (p.p.m.) of the assigned amino acid residues resonances for p(1-11) recorded at 303 K

Residue	Chemical shift (p.p.m.)			
	NH	αH	βH	Others
Pro ¹	8.55 9.22 (min)	4.16	2.24, 1.87	γCH_2 1.87 δCH_2 3.16, 3.19
Asp ²	8.84	4.68	2.74, 2.52	
Thr ³	7.66	4.16	3.93	γCH_3 0.97 OH 4.81
Arg ⁴	7.90 7.45 (min)	4.47	1.69, 1.50	γCH_2 1.50 δCH_2 3.06 NH 7.16, 6.97
Pro ⁵		4.32	2.00, 1.77	γCH_2 1.80 δCH_2 3.60, 3.32
Ala ⁶	8.06	4.43	1.16	
Pro ⁷		4.26	1.92, 1.81	γCH_2 1.80 δCH_2 3.53
Gly ⁸	8.16	3.69		
Ser ⁹	7.74	4.35	3.64, 3.48	OH 5.08
Thr ¹⁰	7.66	4.19	4.05	γCH_3 1.03 OH 4.74
Ala ¹¹	7.90	4.16	1.24

Fig. 4. Temperature-dependence of the amide-proton-resonance chemical shifts for the peptide p(1-11)

The effect of temperature on the resonance position of the amide protons of the peptide p(1-11) dissolved in $[^2\text{H}_6]\text{dimethyl sulphoxide}$ was examined. $\circ$, D^2 (these capital letters with superscripts have the same meaning as in Fig. 2) (0.00287 p.p.m./K); $\triangle$, G^8 (0.00514 p.p.m./K); $\square$, A^6 (0.00501 p.p.m./K); $\bullet$, R^4/A^{11} (0.00456 p.p.m./K); $\blacksquare$, S^9 (0.00364 p.p.m./K); $\blacktriangle$, T^{3+10} (0.00572 p.p.m./K).

Table 3. Through-space connections between the amide protons and the amide and side-chain proton for the peptide p(1-11)

Residue	ROESY cross peak to resonance for:
Asp ²	Thr ⁴ NH
Thr ³	Asp ² NH, Arg ⁴ NH
Arg ⁴	Thr ³ NH, Pro ⁵ $\delta\text{H}_1, \delta\text{H}_2$
Gly ⁸	Ser ⁹ NH

the other protons in this system are of a similar value to the 0.006 p.p.m./K observed for the simple amide *N*-methylacetamide in dimethyl sulphoxide (Millett & Raftery, 1972).

Through-space connections in the ROESY experiment give a further indication of the elements of secondary structure that are present in this peptide. In addition to amide-to- α -proton connectivities along the backbone of the peptide, a number of amide-to-amide and amide-to-side-chain proton connections are observed; these are listed in Table 3. The pattern of connectivity between the amide protons of Asp-Thr-Arg is suggestive of a type I β -turn being present (Wuthrich *et al.*, 1984), as depicted in Fig. 5. The lack of a through-space connection between the α -proton of Asp² and the amide proton of Arg⁴ may be due to the relatively longer distance between these protons [theoretical 0.36 nm (3.6 Å)] compared with the distances from the amide-to-amide protons in the type I β -turn [theoretical Asp²-Thr³, 0.26 nm (2.6 Å); Thr³-Arg⁴, 0.24 nm (2.4 Å)] (Wuthrich *et al.*, 1984), which only give rise to weak ROESY cross peaks. The presence of the through-space connection between the amide proton of Arg⁴ and the δ -protons of Pro⁵ further allows us to identify the conformation of the amide bond between these two amino acids to be *trans*. This effectively extends the turn over the Asp-Thr-Arg-Pro region, which overlaps with the known epitopes for the antibodies C595 and NCRC-11 (Price *et al.*, 1990b). The through-space connection between the amide protons of Gly⁸ and Ser⁹, combined with the rate of exchange of the amide proton of Ser⁹, suggest the presence of a second turn; however, further evidence is needed before this can be substantiated.

Fig. 5. Secondary structure present in peptide p(1-11)

The proposed extended type I β -turn present across the Pro-Asp-Thr-Arg-Pro backbone region of peptide p(1-11) is depicted. The amide-to-amide and amide-to-side-chain through-space connections are also shown ($\bullet$, H; $\circ$, O; $\square$, C).

Indications of there being a number of conformations present in this peptide can be observed directly. Fig. 2 shows the presence of two resonances for the amino proton of Pro¹, with the major conformer resonating at 8.55 p.p.m. Exchange between the two forms has been confirmed via the transfer of magnetization in one-dimensional transfer-of-saturation experiments. As this is the *N*-terminus of the peptide, this interconversion cannot be accounted for by *cis-trans* isomerization. The amide proton of Arg⁴ is also seen to resonate at two different frequencies; the major populated amide resonance has been assigned above to the prolyl *trans* isomer. The exchange of the two conformations of Arg⁴ has been investigated; one-dimensional experiments failed to identify any exchange between the forms giving rise to these resonances. Although at present an impurity cannot be completely ruled out, it would appear unlikely that it would manifest itself with degenerate resonances in all cases except the amide proton of Arg⁴. As the minor populated amide proton resonance for Arg⁴ has no through-space connections to the δ -protons of Pro⁵ it is probable that this represents the *cis* isomer. A number of cross peaks for the threonine residues in the HOHAHA experiment (results not shown) also suggest that these residues are to be found in a number of magnetically inequivalent environments.

Previous studies on the 20-amino-acid-peptide (I) failed to find any evidence for the temperature-dependence of the resonance positions for the high-field methyl signals in both trifluoroethanol and water, suggesting a lack of secondary structure (Price *et al.*, 1990b). This discrepancy with current results may be accounted for by the relative lack of sensitivity of the previous technique when compared with a full one- and two-dimensional n.m.r. characterization of the peptide. Since the peptide is probably present as a mixture of interconverting conformations, it would appear that the relative populations of the conformers are averaged such as to weight the n.m.r. parameters in the manner observed. The combined observations of the through-space connections, and the resonance frequency and solvent-exchange rate of the Asp² amide proton, suggest that presence of an extended type I β -turn in the average conformation of peptide p(1-11) when it is dissolved in dimethyl sulphoxide. This β -turn, shown in Fig. 5, extends into the region that has been previously identified to contain the epitopes for the antibodies C595 and NCRC-11 and may suggest that this structural feature is used by the antibodies to recognize the antigenic peptides.

The support of a Nuffield Foundation New Science Lecturers Award is gratefully acknowledged. Dr. W. C. Chan and Mr. P. Teesdale-Spittle are thanked for the preparation and purification of the peptide used in this study. Dr. M. R. Price is thanked for useful discussions and his support in this project. The experiments on the Bruker AM500 spectrometer were undertaken at the Science and Engineering Research Council Biological NMR Centre at Leicester University. Dr. L. Y. Lian is gratefully acknowledged for her help and advice in this study.

REFERENCES

- Bax, A. & Davis, D. G. (1985) *J. Magn. Reson.* **63**, 207–213
Braunschweiler, L. R. & Ernst, R. R. (1983) *J. Magn. Reson.* **53**, 521–528
Dyson H. J., Cross, K. J., Ostresh, J., Houghten, R. A., Wilson, I. A., Wright, P. E. & Lerner R. A. (1986) in *Synthetic Peptides as Antigens* (Ciba Found. Symp. **119**), pp. 58–75, John Wiley and Sons, Chichester
Dyson, H. J., Lerner, R. A. & Wright, P. E. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 305–324
Gendler, S., Taylor-Papadimitriou, J., Duhig, T., Rothbard, J. & Burchell, J. (1988) *J. Biol. Chem.* **263**, 12820–12823
Hammond, S. J., Birdsall, B., Feeney, J., Searle, M. S., Roberts, G. C. K., & Chung, H. T. A. (1987) *Biochemistry* **26**, 8585–8590
Millett, M. & Raftery, M. A. (1972) *Biochemistry* **11**, 1639–1643
Price, M. R. (1988) *Eur. J. Cancer Clin. Oncol.* **24**, 1799–1804
Price, M. R., Pugh, J. A., Hudecz, F., Griffiths, W., Jacobs, E., Clarke, A. J., Chan, W. C. & Baldwin, R. W. (1990a) *Br. J. Cancer*, in the press
Price, M. R., Hudecz, F., O'Sullivan, C., Baldwin, R. W., Edwards, P. M. & Tendler, S. J. B. (1990b) *Mol. Immunol.*, in the press
Wright, P. E., Dyson, H. J. & Lerner, R. A. (1988) *Biochemistry* **27**, 7167–7175
Wuthrich, K. (1986) *NMR of Proteins and Nucleic acids*, John Wiley and Sons, New York
Wuthrich, K., Billeter, M. & Braun, W. (1984) *J. Mol. Biol.* **180**, 715–740

Received 22 September 1989/5 December 1989; accepted 7 December 1989